



## Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP

Fabian Kirstein, Benjamin Dittwald, Simon Dutkowski, Yury Glikman, Sonja Schimmler, Manfred Hauswirth

### ► To cite this version:

Fabian Kirstein, Benjamin Dittwald, Simon Dutkowski, Yury Glikman, Sonja Schimmler, et al.. Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP. 18th International Conference on Electronic Government (EGOV), Sep 2019, San Benedetto del Tronto, Italy. pp.192-204, 10.1007/978-3-030-27325-5\_15 . hal-02445810

**HAL Id: hal-02445810**

**<https://inria.hal.science/hal-02445810>**

Submitted on 20 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP

Fabian Kirstein<sup>1,2</sup> ✉, Benjamin Dittwald<sup>1</sup>, Simon Dutkowski<sup>1</sup>, Yury Glikman<sup>1</sup>,  
Sonja Schimmler<sup>1,2</sup>, and Manfred Hauswirth<sup>1,2,3</sup>

<sup>1</sup> Fraunhofer FOKUS, Berlin, Germany

<sup>2</sup> Weizenbaum Institute for the Networked Society, Berlin, Germany

<sup>3</sup> TU Berlin, Open Distributed Systems, Berlin, Germany  
`{firstname.lastname}@fokus.fraunhofer.de`

**Abstract.** The European Data Portal (EDP) is a central access point for metadata of Open Data published by public authorities in Europe and acquires data from more than 70 national data providers. The platform is a starting point in adopting the Linked Data specification DCAT-AP, aiming to increase interoperability and accessibility of Open Data. In this paper, we present the design of the central data management components of the platform, responsible for metadata storage, data harvesting and quality assessment. The core component is based on CKAN, which is extended by the support for native Linked Data replication to a triplestore to ensure legacy compatibility and the support for DCAT-AP. Regular data harvesting and the creation of detailed quality reports are performed by custom components addressing the requirements of DCAT-AP. The EDP is well on track to become the core platform for European Open Data and fostered the acceptance of DCAT-AP. Our platform is available here: <https://www.europeandataportal.eu>

**Keywords:** Open Data · Linked Data · CKAN · DCAT-AP · Data and Information Quality

**Track:** Open Data: Social and Technical Aspects

## 1 Introduction

Open Data is a driver for transparency and innovation. Freely available machine-readable data can help to foster participation and may create novel business models [21]. Typical Open Data is weather data, geographical data, traffic data, statistics, publications, protocols, laws and ordinances. The publication of Open Data is mainly conducted by public administrations and organisations, but a growing number of private companies have begun to initiate Open Data projects as well. If data is available as Open Data, it can be used, processed, refined and distributed by everyone at any time without mandatory registration, without restrictions and free of charge. The most typical channel of distribution of Open Data is through a Web portal. As of today, more than 2,600 portals exist [11]. In order to encourage reuse and application of Open Data, a common standard for storing and managing metadata is advisable. Especially, standardized access via an Application Programming Interface (API) is indispensable. Since 2013, a specification for describing public sector datasets in Europe, called DCAT Application profile for data portals in Europe (DCAT-AP) is developed by order of the European Commission (EC). The profile is based on Linked Data principles and the Resource Description Framework (RDF) vocabulary Data Catalogue Vocabulary (DCAT). It is designed to increase interoperability and allows the user to search for Open Data across multiple portals. The standard is constantly refined and currently published in version 1.2 [5].

### 1.1 The European Data Portal (EDP)

In November 2015, the EC launched the European Data Portal (EDP), which makes all metadata of Open Data published by public authorities of the European Union (EU) available in one portal [7]. As of February 2019, the EDP lists close to 900.000 datasets, in total consisting of about 60 million RDF triples, harvested from 77 data providers.<sup>4</sup> The EDP is Europe’s Linked Data-enabled one-stop-shop for open public sector information. It is not limited to a metadata registry, but forms an entire ecosystem for fostering the manifestation, reuse and quality improvement of Open Data. The platform pioneered in adopting the DCAT-AP specification and represents its first reference implementation. The core metadata properties are available in all 24 official languages of the EU. Where translations are not provided by the original data provider, a machine-translation service is employed.

The design and implementation of the EDP posed some extensive challenges: (i) The user interface and API, on the one hand, had to be compliant with already established non-Linked Data standards of Open Data publishing in order to meet the expectations of Open Data users. The metadata, on the other hand, had to be stored in a native RDF data model, complying with the new DCAT-AP specifications. Hence, Linked Open Data (LOD) had been required, which enables the access to the metadata via a SPARQL endpoint. Therefore, a **metadata**

---

<sup>4</sup> <https://www.europeandataportal.eu/data/#/catalogues>

**registry**, bridging these two concepts satisfactorily had been required for the EDP. (ii) Metadata from all European national Open Data portals had to be retrieved, harmonized and made available. Updates in the source metadata had to be reflected on the EDP without delay. Existing Open Data fetching tools did not fit the diversity and volume of the data providers for the EDP. Therefore, a suitable **harvesting** mechanism needed to be developed for the EDP. (iii) The completeness and compliance of metadata are key factors for a successful Open Data platform. Such metrics are rarely accessible or even reviewed. Therefore, a central aspect of the EDP had been the provision of **metadata quality** reports. In this paper, we present the concept, the implementation, and our lessons learned from the EDP infrastructure with the key challenge to comply to the DCAT-AP specification. The paper is focused on the central components, forming the so-called EDP data segment:<sup>5</sup> the *Metadata Registry*, the *Harvester* and the *Metadata Quality Assurance*. After an overview of related work and established Open Data standards (Section 2), the high-level design is described (Section 3). In Section 4, solution statements for each central component (i.e., service) are given. The application of DCAT-AP, namely the linked data management, is then highlighted in Section 5. Finally, the impact of the presented work is evaluated (Section 6) and directions for future research are presented (Section 7).

## 2 Related Work

DCAT is a widely adopted and popular standard for describing datasets and establishing interoperability between data catalogues. DCAT-AP is a Linked Data extension of DCAT which adds metadata fields and mandatory ranges for specific properties [5]. These ranges are mostly provided as a Simple Knowledge Organization System (SKOS) controlled vocabulary, provided by the EC Publications Office. For instance, properties like language, spatial information or MIME type can be harmonised by applying the provided vocabulary [15]. The popularity of DCAT-AP is increasing and country-specific extensions have been published. E.g., the German IT Planning Council established DCAT-AP.de as the official exchange standard for open governmental data in Germany [8].

A lot of work has been invested in developing tools for making Open Data and Linked Data publicly available and easily accessible. The Open Source solution Comprehensive Knowledge Archive Network (CKAN) is a basic web application for building data catalogues, particularly for Open Data. It is the de-facto standard in the public sector, but is also applied by private companies. CKAN provides many features for mapping the process of publishing data catalogues. A comprehensive range of plug-ins is available. The CKAN API is extensively documented and provides a comprehensive way to retrieve the metadata of the data catalogue [4].

There exist several hybrid approaches, where the Linked Data interface is an

---

<sup>5</sup> The entire EDP includes further tools, which are out of scope, e.g., a content management system, tools for visualisation and preview, a proxy-server and a load balancer.

additional layer on top of traditional data structures. An official plug-in for extending CKAN with a DCAT-AP interface is available.<sup>6</sup> However, it only maps the existing data structures to an RDF serialisation and does not provide native Linked Data capabilities. A proprietary and closed source alternative to CKAN is OpenDataSoft, which is also used by a variety of institutions for implementing Open Data platforms [13]. It focuses on interaction and visualization through automated API generation and has only limited support for DCAT-AP. An interoperability mode for mapping the default schema to DCAT-AP is available on demand [12]. The open access repository software DSpace follows a more elaborate approach [16]. A converter is available, which dynamically translates relational metadata into native RDF metadata. The converter stores the generated triples into a triplestore, and hence, offers the metadata via a SPARQL endpoint [6]. A very similar approach is followed by Wikidata, a community-driven knowledge base by the Wikimedia Foundation. Wikidata implements a custom data structure for storing statements about identifiable items, very similar to the concept of RDF. This data is periodically converted to native RDF and stored in a triplestore, whose endpoint is publicly accessible [17].

As an alternative to the above mentioned hybrid approaches, systems can be built upon native Linked Data. The W3C recommendation for Linked Data Platforms (LDPs) defines a low-level specification for managing Linked Data resources on the web. It is based on HTTP methods and defines guidelines for representation formats, collision detection and vocabulary reuse [18]. Several implementations of LDP exist, like OpenLink Virtuoso or Apache Marmotta [20]. The latter builds upon a straightforward native application of RDF with pluggable triplestores and targets organisations that want to publish Linked Data [1]. Virtuoso is a full-fledged and mature database management solution, supporting and combining multiple paradigms for storing data in a unique system. Foremost, it can be used as a highly scalable and versatile triplestore [14]. Klímek et al. [9] present a first system, which is entirely built upon native DCAT-AP and is used in the Czech National Open Data Catalog<sup>7</sup>. An interactive pipeline process is applied for harvesting DCAT-AP from official institutions and storing the metadata in both, a triplestore and an Apache Solr [2] search index.

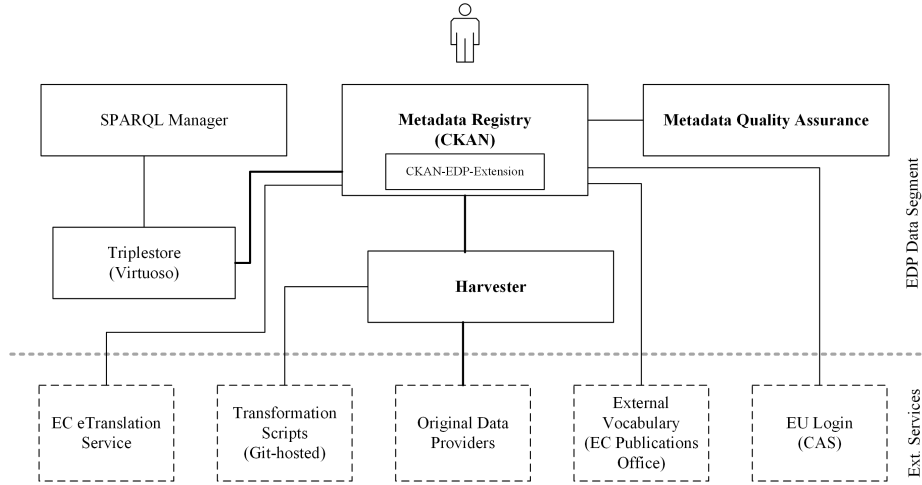
In order to provide a practical and flexible solution, we also follow a hybrid approach, where all metadata is represented in both formats.

### 3 High-Level Design

The central objective of the design of the EDP is to address the requirements of the DCAT-AP-compliant metadata storage, the data acquisition and the quality assurance in a practical and scalable way. The general design follows a service-oriented approach, with a strict separation of concerns. This ensures high scalability, since every service can run independently on a separate machine. In addition, the services are designed statelessly, whenever possible, hence, allowing

<sup>6</sup> <https://github.com/ckan/ckanext-dcat>

<sup>7</sup> <https://data.gov.cz>



**Fig. 1.** Overview of the components of the EDP data segment

for a replication on multiple machines, if necessary. Fig. 1 illustrates the interactions and deployment of the internal and external components. All services communicate via RESTful APIs with each other. The distributed architecture requires a central authentication service for securing restricted operations of the platform (e.g., for creating new datasets). The authentication service of the EC (EU Login<sup>8</sup>) is integrated for that purpose. It implements the single sign-on protocol Central Authentication Service (CAS).

To enable backward compatibility within the existing European Open Data ecosystem, CKAN was used as a basis. This way, we ensure the provisioning of mature elementary features and the compliance with established methodologies and interfaces. Additional and modified functionalities of CKAN are implemented based on its rich extension interfaces, resulting in the *CKAN EDP extension*.

The central service of the EDP data segment is our **Metadata Registry**, which includes some features that concern the storage and management of metadata. Here, native DCAT-AP is integrated with a proven replication approach (see Section 2), where all metadata is additionally stored in a triplestore. As an appropriate solution for the diverse Open Data acquisition and transformation tasks, our **Harvester** was implemented as a second service of the EDP data segment. Based on custom transformation scripts, it is responsible for fetching the metadata and for converting it into the target data formats. As a third service of the EDP data segment, our **Metadata Quality Assurance (MQA)** service is continuously retrieving the metadata from the registry and is validating it against the target schema DCAT-AP. The validation results are summarized and accessible to the data provider.

<sup>8</sup> <https://webgate.ec.europa.eu/cas>

## 4 Service Design

In the following, the three central services of the EDP data segment, which were overviewed in the last section, are discussed along with our approaches for solving the challenges we faced.

### 4.1 Metadata Registry

The Metadata Registry acts as the central data management unit for the EDP and the primary access point for users of the metadata. It adopts its core technology stack from the underlying CKAN, which is implemented in Python and which uses PostgreSQL as storage and Apache Solr for search. Fig. 2 shows the main search page served by the Metadata Registry.

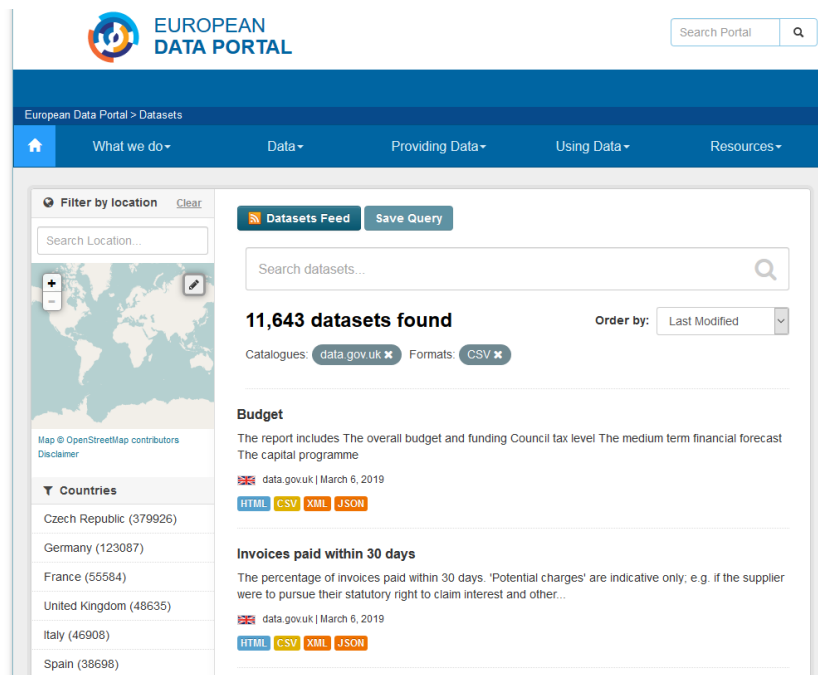


Fig. 2. Main search page for datasets

The CKAN EDP extension<sup>9</sup> introduces multiple features into the core system, where the most significant one is the management of DCAT-AP (see Section 5). Significant modifications concern the default data structure of CKAN, the CKAN Data Schema (CDS), which is extended. This includes support for textual metadata and other meta information in multiple languages. The underlying

<sup>9</sup> <https://gitlab.com/european-data-portal/ckanext-edp>

schema of the search server Solr is adjusted accordingly. Especially, language-aware analysers are introduced, e.g., for stemming. The actual translations are provided by the eTranslation service of the EC. In order to avoid any blocking functionality, the Metadata Registry accumulates certain literal metadata fields to batches and sends them to the eTranslation service. The Metadata Registry is then asynchronously updated with the machine-translated texts retrieved.

## 4.2 Harvester

The Harvester is a standalone service for acquiring metadata from the various source portals, and for transforming it into the DCAT-AP-compliant data structure of the EDP. It is designed to fit the specific needs of the EDP to deal with a wide diversity of source data formats and high update rates. Existing solutions for harvesting, especially the CKAN Harvest<sup>10</sup> extension could not cope well with these requirements. Especially, the development and management of custom adapters were too complex.

The Harvester currently supports 12 different input protocols and serialisation formats, most prominently, CKAN-API, SPARQL, RSS and OpenDataSoft-API, and the DCAT-AP-compliant data structure of the EDP as output format. A harvester represents a link between two repositories, where one acts as source and one as target. Each harvester is defined by transformation rules, which define how the source serialisation is converted into the target one. These rules are defined with simple scripting languages. Currently, we support eXtensible Stylesheet Language Transformations (XSLT) for XML serialisation and JavaScript for JSON serialisation. The scripts are managed by the system and are dynamically loaded.<sup>11</sup> This enables an agile reaction on changes in the payload of a data source. Listing 1 shows an excerpt of a transformation script. For configuring harvesters, a user-friendly web frontend is available. A harvester has a schedule, which can be configured individually (e.g., daily, weekly etc.). In order to avoid unnecessary update operations, a hash value for each source dataset is calculated and stored in the Metadata Registry. An update is only triggered, when the source data, and therefore the hash has changed.

---

```
if (input.organization) {  
    output.publisher = {};  
    output.publisher.type = "http://xmlns.com/foaf/0.1/Organization";  
    output.publisher.name = input.organization.title;  
}
```

---

**Listing 1.** Excerpt of a JavaScript transformation script

## 4.3 Metadata Quality Assurance (MQA)

The MQA<sup>12</sup> is a service, which periodically executes quality checks of the metadata stored in the Metadata Registry. Existing tools for Open Data quality

<sup>10</sup> <https://github.com/ckan/ckanext-harvest/>

<sup>11</sup> <https://gitlab.com/european-data-portal/transformation-scripts>

<sup>12</sup> <https://www.europeandataportal.eu/mqa>



assessment typically work on a less granular level and are updated infrequently, for example the Open Data Monitor.<sup>13</sup>

The validation is conducted on two levels: (i) First, the formal correctness of the metadata is checked. This is done by validating each set of metadata against a self-defined JSON Schema. This schema includes the constraints and specifications of DCAT-AP. It was chosen to perform the validation against the extended CKAN-API due to the maturity of JSON-based validation. Validation tools for RDF, like SHACL [19], had not been advanced enough at the time of development. The check provides a detailed report of schema validations, e.g., missing mandatory properties or wrong data types. (ii) Second, the actual content of specific metadata properties, which are shown in Table 1, is checked. The results are aggregated and visualised as illustrated in Fig. 3.

**Table 1.** Validation characteristics of the MQA

Feature	Description
<b>Accessible Distributions</b>	A HTTP GET request is performed on all distributions in order to determine their accessibility.
<b>Error Status Codes</b>	If a distribution is not accessible, the HTTP error code is logged and reported.
<b>Ratio Machine-Readable Datasets</b>	The ratio of machine-readable and non-machine-readable distributions is calculated. The determination is based on a list published by the Open Data Monitor. <sup>14</sup>
<b>Most Used Formats</b>	The most used data formats are presented.
<b>Ratio Known to Unknown Licences</b>	The ratio of known and unknown licenses is calculated. Therefore, the used licences are validated against a comprehensive list of Open Data licences. <sup>15</sup>
<b>Most Used Licences</b>	The most used licences are presented.

In addition, the MQA computes a list of datasets that are similar to a certain dataset based on their metadata. The feature uses a locality-sensitive hashing (LSH) algorithm [10] and the result is presented in the Metadata Registry. The MQA performs a full validation of the entire data pool of the EDP once every month. The entire process is resource-intensive due to many external dependencies, e.g., resolving the URLs is time consuming.

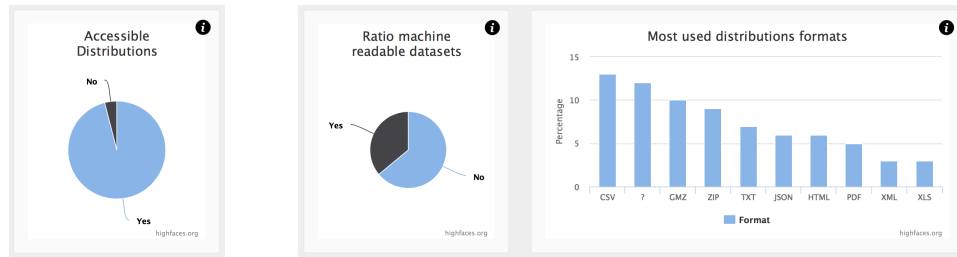
## 5 Linked Data Management

The support of the DCAT-AP specification is a core feature of the EDP and handled by the Metadata Repository. The majority of the data providers do

<sup>13</sup> <https://opendatamonitor.eu>

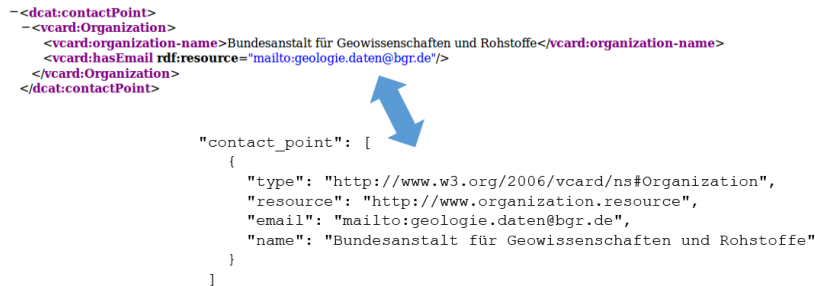
<sup>14</sup> [https://github.com/opendatamonitor/odm.restapi/blob/master/odmapi/def\\_formatLists.py](https://github.com/opendatamonitor/odm.restapi/blob/master/odmapi/def_formatLists.py)

<sup>15</sup> <https://www.europeandataportal.eu/en/content/show-license>



**Fig. 3.** MQA quality report

not serve DCAT-AP-compliant metadata or even Linked Data. In addition, the underlying CKAN of the Metadata Registry does not support native Linked Data (see Section 2). Therefore, a triplestore is introduced as additional database and replication layer. Virtuoso is used here, due to its maturity and LDP-compliance. Hence, the source metadata needs to be represented in both formats, DCAT-AP and CKAN-JSON. Both serialisations have to be idem-potent, i.e., bi-directional conversion is possible without any losses. Therefore, a virtual data format for creating and managing the metadata is required. The CDS is extended in order to match the DCAT-AP data schema. All DCAT-AP properties and structures are mapped to the CDS correspondents and additional properties are added. Fig. 4 shows an exemplary mapping for the property `contactPoint`.<sup>16</sup> The use of the extended CDS as the common data structure ensures compatibility with established practices.



**Fig. 4.** Mapping of DCAT-AP to CDS and vice versa

The Metadata Registry is exposing a compliant interface for both, the API and the Web frontend. When a new dataset is created, it is converted to native RDF and stored in the triplestore. The Python library `rdflib` is applied for that purpose. For each property, detailed creation rules are provided, en-

<sup>16</sup> Full mapping of all properties: <https://gitlab.com/european-data-portal/ckanext-edp/wikis/schema>

ensuring the creation of rich DCAT-AP, following best practices for Linked Data. The creation of persistent URIs follows established practices [3]. For instance, the URI scheme `http://europeandataportal.eu/set/data/[name]` is used for datasets. The Linked Data representations of the entire registry can be accessed via a SPARQL endpoint. The integrated SPARQL Manager allows for the interactive creation and management of arbitrary queries. To release the full potential of the underlying LOD, the Metadata Registry supports content negation, allowing clients to retrieve every dataset in different RDF serialisation formats, by adding a trailing format indicator, e.g., `.rdf` or `.ttl`. Finally, the Metadata Registry is resolving properties (described with URIs of known vocabularies) to human-readable representations within the frontend. Mostly, the controlled vocabulary from the EC Publications Office is used here.

## 6 Evaluation

The EDP is employed in production, and is publicly accessible since November 2015. As of February 2019, it offers close to 900.000 datasets, from 35 countries in Europe and beyond, provided by 77 data providers. In total, about 60 million RDF triples are accessible. Its use and acceptance is constantly monitored since its launch. The web analytics tool Matomo<sup>17</sup> is integrated in all services for that purpose. Since its launch, the EDP had more than 870.000 unique visitors. Detailed statistics are aggregated on a quarterly basis. As an example, Table 2 shows an extraction of the data for Q2 to Q4 2018. Overall, the statistics show a productive adoption and stable service of the EDP. The numbers for the Page Views Data Segment are an indicator for the use of the components and services, which have been described in Section 4. The numbers for the SPARQL Manager are an indicator for the use of the native Linked Data, which has been described in Section 5.

**Table 2.** Extract of the use statistics of the EDP

	Q2 2018	Q3 2018	Q4 2018
<b>Page Views</b>	341.891	296.899	300.260
<b>Page Views Data Segment</b>	257.183	227.221	220.204
<b>SPARQL Manager</b>	2.954	2.620	2.635
<b>Total Visits</b>	114.570	102.454	99.772
<b>Unique Visitors</b>	105.989	94.852	91.736
<b>Daily Average</b>	1.259	1.114	1.084
<b>Returning Visitors</b>	19%	19%	20%

The acceptance of DCAT-AP has increased significantly since the release of the EDP. In 2015, only one data provider served DCAT-AP, whereas now, as of February 2019, 13 providers follow the specification. The transparent communication, the accessible reports of the MQA and the high visibility of the EDP

<sup>17</sup> formerly Piwik

helped fostering the awareness of the specification. However, the vast diversity of employed data structures, formats and vocabularies used by the different data providers impede a homogeneous presentation of the available data in the EDP. This has an impact on the usability and lowers the much higher potential of Open Data. Furthermore, the underlying platform CKAN is a great choice for building regional and national Open Data portals, but does not easily perform smoothly with the amount of data hosted by the EDP. It was necessary to restrict the update frequency to ensure stability and availability.

## 7 Conclusion and Outlook

In this paper, we have presented the design and implementation of the European Data Portal with focus on the EDP data segment and the adoption of the DCAT-AP Linked Data specification. With the EDP, we provide a comprehensive platform for acquiring, presenting and validating Open Data from all over the EU. It has established itself as a well-known one-stop-shop for Open Data and an advocate for the Linked Open Data movement. The presented approach successfully combines two concepts of representing and serving Open Data: the established data structure and API of CKAN and Linked Data via SPARQL. Currently, more and more national and regional data providers adopt the DCAT-AP standard, which the EDP already supports. In parallel, the importance of native and simple LOD increases. As a consequence, the replication of the metadata in two distinct data storages, which is complex, should be adapted. In addition, a transition from restricted, flat data structures to Linked Open Data is on-going. Furthermore, it can be expected that the overall amount of Open Data in Europe will increase. Thus, a stronger focus on high-performance solutions and support for faster harvesting updates will be necessary. Therefore, we currently adapt the presented architecture to accommodate these future requirements. A revised and improved version of the EDP is under active development, considering the learnings from the first EDP version described in this paper.

## Acknowledgments

This work has been funded by the European Commission, Directorate-General Communications Networks, Content and Technology under service contract for SMART 2014/1072 "Deployment of an EU Open Data core platform: implementation of the pan-European Open Data portal and related services" and by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII111 ("Deutsches Internet-Institut"). We would like to thank the European Commission for granting us access to the usage statistics. We would also like to thank the entire consortium of the European Data Portal for making this work possible.

## References

1. Apache Software Foundation: Apache Marmotta, <http://marmotta.apache.org/>
2. Apache Software Foundation: Apache Solr, <http://lucene.apache.org/solr/>
3. Archer, P., Goedertier, S., Loutas, N.: D7.1.3 - Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC (Dec 2012), <https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>
4. CKAN Association: CKAN, <https://ckan.org/>
5. Dragan, A.: DCAT Application Profile for data portals in Europe (Nov 2018), [https://joinup.ec.europa.eu/sites/default/files/distribution/access\\_url/2018-11/014bde52-eb3c-4060-8c3c-fcd0dfc07a8a/DCAT\\_AP\\_1.2.pdf](https://joinup.ec.europa.eu/sites/default/files/distribution/access_url/2018-11/014bde52-eb3c-4060-8c3c-fcd0dfc07a8a/DCAT_AP_1.2.pdf)
6. DuraSpace Wiki: Linked (Open) Data, <https://wiki.duraspace.org/display/DSDOC6x/Linked+%28Open%29+Data>, (Accessed: 11.03.2019)
7. European Data Portal: The European Data Portal: Opening up Europe's public data, [https://www.europeandataportal.eu/sites/default/files/edp\\_factsheet\\_what\\_is\\_edp\\_project\\_online.pdf](https://www.europeandataportal.eu/sites/default/files/edp_factsheet_what_is_edp_project_online.pdf), (Accessed: 11.03.2019)
8. [init] AG und SID Sachsen: DCAT-AP.de Spezifikation, <https://www.dcat-ap.de/def/dcatde/1.0.1/spec/specification.pdf>, (Accessed: 11.03.2019)
9. Klímek, J., Skoda, P.: LinkedPipes DCAT-AP Viewer: A Native DCAT-AP Data Catalog. In: International Semantic Web Conference (2018), <https://pdfs.semanticscholar.org/28ab/7bcd1b5db660ac280e426556e96d599daed.pdf>
10. Oliver, J., Cheng, C., Chen, Y.: TLSH – A Locality Sensitive Hash. In: 2013 Fourth Cybercrime and Trustworthy Computing Workshop. pp. 7–13 (Nov 2013). <https://doi.org/10.1109/CTC.2013.9>
11. OpenDataSoft: A Comprehensive List of 2600+ Open Data Portals around the World, <https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/>, (Accessed: 11.03.2019)
12. OpenDataSoft: Interoperability Metadata, [https://help.opendatasoft.com/platform/en/publishing\\_data/06\\_configuring\\_metadata/interoperability\\_metadata.html](https://help.opendatasoft.com/platform/en/publishing_data/06_configuring_metadata/interoperability_metadata.html), (Accessed: 11.03.2019)
13. OpenDataSoft: Open Data Solution, <https://www.opendatasoft.com/solutions/open-data/>, (Accessed: 11.03.2019)
14. OpenLink Software: About OpenLink Virtuoso, <https://virtuoso.openlinksw.com/>, (Accessed: 11.03.2019)
15. Publications Office of the EU: Authority tables, <https://publications.europa.eu/en/web/eu-vocabularies/authority-tables>, (Accessed: 11.03.2019)
16. Smith, M., Barton, M., Bass, M., Branschovsky, M., McClellan, G., Stuve, D., Tansley, R., Walker, J.H.: DSpace: An Open Source Dynamic Digital Repository (Jan 2003). <https://doi.org/DOI:10.1045/january2003-smith>
17. Vrandečić, D., Krötzsch, M.: Wikidata: A Free Collaborative Knowledgebase. Commun. ACM **57**(10), 78–85 (Sep 2014). <https://doi.org/10.1145/2629489>
18. W3C: Linked Data Platform 1.0, <https://www.w3.org/TR/ldp/>
19. W3C: Shapes Constraint Language (SHACL), <https://www.w3.org/TR/shacl/>
20. W3C Wiki: LDP Implementations, [https://www.w3.org/wiki/LDP\\_Implementations](https://www.w3.org/wiki/LDP_Implementations), (Accessed: 11.03.2019)
21. Zaki, M., Feldmann, N., Neely, A., Hartmann, P.M.: Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. International Journal of Operations & Production Management **36**(10), 1382–1406 (Sep 2016). <https://doi.org/10.1108/IJOPM-02-2014-0098>